カーネル法による因果ネットワークの学習

情報・システム研究機構　統計数理研究所　福水　健次[1]

本稿は、[8] にしたがって、正定値カーネルを用いた確率変数の独立性、条件付独立性の尺度を用いて、与えられたデータから変数間の因果ネットワークを推定する方法の概要を述べる。

1　正定値カーネルによる独立性・条件付独立性尺度

まず、正定値カーネルの基本事項をまとめておく。集合 Ω に対し、$k:\Omega\times\Omega\to\mathbb{R}$ が Ω 上の正定値カーネルであるとは、対称性 $k(x,y)=k(y,x)$ を満たし、かつ任意の n 個の点 $x_1,\ldots,x_n\in\Omega$ と 実数 $c_1,\ldots,c_n$ に対し、

$$\sum_{i,j=1}^{n} c_i c_j k(x_i,x_j)\geq 0$$

が成り立つことをいう。

Ω 上の正定値カーネル k に対し、集合 Ω 上の実関数からなる（実）ヒルベルト空間 $\mathcal{H}$ が存在し、以下の 2 つの性質を満たす。

(i) 任意の $x\in\Omega$ に対して $k(\cdot,x)\in\mathcal{H}$ であり、$\{k(\cdot,x)\in\mathcal{H}\mid x\in\Omega\}$ の張る線形空間は $\mathcal{H}$ の中で稠密である。

(ii) 任意の $f\in\mathcal{H}$ と $x\in\Omega$ に対し、再生性

$$\langle f,k(\cdot,x)\rangle_{\mathcal{H}}=f(x)$$

が成り立つ。ここで $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ は $\mathcal{H}$ の内積を表す。

このようなヒルベルト空間のことを（k が定める）再生核ヒルベルト空間といい、$(\mathcal{H},k)$ であらわす。(ii) の再生性は再生核ヒルベルト空間をデータ解析に応用する上で重要な性質であり、内積計算やデータによる推定量の構成を容易にする。

ユークリッド空間 $\mathbb{R}^m$ 上の正定値カーネルの代表的な例は、通常の内積 $k(x_1,x_2)=x_1^Tx_2$ のほかに、ガウス RBF(Radial Basis Function) カーネル

$$k_\sigma^G(x_1,x_2)=\exp\Bigl(-\frac{\|x_1-x_2\|^2}{2\sigma^2}\Bigr)$$

($\sigma>0$) などがある。

---

[1]E-mail: fukumizu@ism.ac.jp

次に、正定値カーネルを変数の依存性解析に用いる方法を述べる。いま可測集合 $\mathcal{X}, \mathcal{Y}$ 上に可測正定値カーネル $k_\mathcal{X}$,$k_\mathcal{Y}$ が与えられており、対応する再生核ヒルベルト空間を $\mathcal{H}_\mathcal{X}$, $\mathcal{H}_\mathcal{Y}$ とかく。また、(X,Y) を $\mathcal{X}\times\mathcal{Y}$ に値をとる確率変数とする。このとき、

$$\Phi_\mathcal{X} : \mathcal{X} \to \mathcal{H}, \qquad \Phi(x) = k_\mathcal{X}(x, \cdot)$$

は $\mathcal{H}_\mathcal{X}$ に値を持つ確率変数 $\Phi_\mathcal{X}(X)$ を定める。確率変数 $\Phi_\mathcal{X}(X)$ は、正定値カーネルの非線形性に従って、X の高次モーメントの情報を保有している。また、$\Phi_\mathcal{Y}(Y)$ も同様に定義できる。

以下に見るように、確率ベクトル $\Phi_\mathcal{X}(X)$ などの基本的な統計量を考えると、もとの確率変数の高次モーメントを扱うことができ、独立性や条件付独立性を特徴付けることができる。以下文献 [1] [3] にしたがってこれを述べる。

$E[k_\mathcal{X}(X,X)], E[k_\mathcal{Y}(Y,Y)]$ が有限であるとき、$\Phi_\mathcal{X}(X)$ と $\Phi_\mathcal{Y}(Y)$ の共分散を考えることができる。実際、$\mathcal{H}_\mathcal{X}$ から $\mathcal{H}_\mathcal{Y}$ への有界作用素 Σ_{YX} が定まり、

$$\langle g, \Sigma_{YX} f\rangle_{\mathcal{H}_\mathcal{Y}} = \mathrm{Cov}[f(X), g(Y)] = E[f(X)g(Y)] - E[f(X)]E[g(Y)] \tag{1}$$

を満たす。Σ_{YX} を相互共分散作用素と呼ぶ。これは、一般に X と Y の高次相関の情報を持っている。

実は、ユークリッド空間上のガウス RBF カーネルなどを含むあるクラスの正定値カーネル[2]を用いると、Σ_{YX} によって X と Y の独立性を特徴付けることができる。

$$X \perp\!\!\!\perp Y \iff \Sigma_{XY} = O \quad (\text{零作用素}) \tag{2}$$

さらに、第3の確率変数 Z に対しても同様に正定値カーネルを定め、条件付相互共分散作用素を

$$\Sigma_{YX|Z} = \Sigma_{YX} - \Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}$$

により定義する。このとき、拡張された変数 $\tilde{X} = (X,Z), \tilde{Y} = (Y,Z)$ を定めると、条件付独立性が以下のように特徴付けられる。

$$X \perp\!\!\!\perp Y \mid Z \iff \Sigma_{\tilde{Y}\tilde{X}|Z} = O \tag{3}$$

以上の事実により、それぞれの Hilbert-Schmidt ノルム

$$H_{YX} = \|\Sigma_{YX}\|_{HS}^2$$
$$H_{YX|Z} = \|\Sigma_{\tilde{Y}\tilde{X}|Z}\|_{HS}^2$$

を、独立性、条件付独立性の尺度として用いることができる。また、これらの尺度の有限サンプルによる推定量が容易に構成でき、一致性などの統計的性質も知られている。

[2] 文献 [2] では「特徴的なカーネル」として定義している

2 因果推論への応用

受動的に与えられたデータから因果ネットワークを推論する方法として、Inductive Causation ([4]) がある。これは、非巡回有向グラフ (DAG) によって表されるマルコフ性を満たす確率分布からデータが発生したという仮定のもとで、変数間の条件付独立性を判定することにより、もとの DAG の同値類を復元する方法である。基本的なアイデアは、$X \to Y \leftarrow Z$ という有向グラフで表される確率分布は、$X \perp\!\!\!\perp Z$ と $X \not\perp\!\!\!\perp Z|Y$ という条件で特徴付けられるという事実である。このような構造を V-structure と呼び、$X \perp\!\!\!\perp Z$ と $X \not\perp\!\!\!\perp Z|Y$ を判定することにより、矢印の向きを決めることができる。一方、$X \to Y \to Z$ と $X \leftarrow Y \leftarrow Z$ は同一の同時分布を定めるため、データからは区別をすることはできない。すなわち、このような場合には矢印の向きを決めることはできない。したがって、DAG の復元においては、すべての矢印の向きを決定できるわけではない。

Inductive causation のアルゴリズムの概要は以下のとおりである。

1. 変数の部分集合の三つ組みに対して条件付独立性を判定し、無向グラフを作成する。
2. 条件付独立性の情報から、すべての V-structure に対して矢印をつける。
3. ループと新たな V-structure を構成しない限り、無向のエッジに向きをつける。

Inductive Causation の具体的なアルゴリズムとしては PC アルゴリズムが知られている ([5])。PC アルゴリズムを含む多くの方法では、条件付独立性の判定に離散変数かガウス性を仮定しており、一般の連続確率変数に対して適用可能なアルゴリズムは少ない。これは、連続変数に対する条件付独立性のよい検定法があまりないことに原因がある。

そこで我々は、前章で述べた正定値カーネルに基づく条件付独立性尺度を用いて、並べ替え検定を行うことによって、Inductive Causation を実装した。さらに、Inductive Causation の第 1 ステップでの統計的誤判定に対処するため、第 2 ステップで voting による向き付けを行った。その詳しいアルゴリズムや実験結果に関しては、文献 [7] を見ていただきたい。また、正定値カーネルを用いた、さらに発展的な因果推論法の詳細に関しては [6] を参照していただきたい。

謝辞

本研究の一部は、Humboldt 財団フェローシップおよび科研費基盤研究 (C)19500249 の補助を受けた。

参考文献

[1] K. Fukumizu, F. R. Bach, and M. I. Jordan. Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *Journal of Machine Learning Research*, 5:73–99, 2004.

[2] K. Fukumizu, A. Gretton, X. Sun, and B. Schölkopf. Kernel measures of conditional dependence. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 489–496. MIT Press, Cambridge, MA, 2008.

[3] K. Fukumizu, F. R.Bach, and M. I. Jordan. Kernel dimension reduction in regression. Technical Report 715, Department of Statistics, University of California, Berkeley, 2006.

[4] J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, 2000.

[5] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction and Search.* Springer Verlag, 1993.

[6] X. Sun. Causal inference from statistical data. Ph.D. thesis, Universität Karlsruhe, 2008.

[7] X. Sun, D. Janzing, B. Schölkopf, and K. Fukumizu. A kernel-based causal learning algorithm. In *Proceedings of International Conference on Machine Learning*, 2007 to appear.

[8] X. Sun, D. Janzing, B. Schölkopf, K. Fukumizu, and A. Gretton. Learning causal inference via kernel-based statstical dependence measures. *in preparation*, 2007.